

(19)



JAPANESE PATENT OFFICE

PATENT ABSTRACTS OF JAPAN

(11) Publication number: **01244497 A**(43) Date of publication of application: **28.09.89**

(51) Int. Cl.

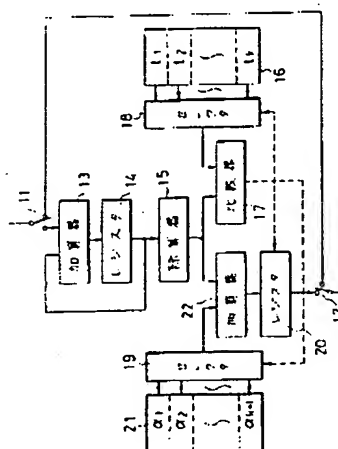
G10L 3/00**G06F 3/16**(21) Application number: **63069791**(71) Applicant: **TOSHIBA CORP.**(22) Date of filing: **25.03.88**(72) Inventor: **SUGI NOBUO**(54) **SOUND SECTION DETECTION CIRCUIT**

COPYRIGHT: (C)1989,JPO&Japio

(57) Abstract:

PURPOSE: To set a pertinent threshold for detecting sound section by adjusting in accordance with the scale of noises so that a sound section can be detected with high accuracy by setting the threshold by adding a constant corresponding to the scale of the noises to the mean value of noise power.

CONSTITUTION: Input signals fetched through a switch 11 are added to signals stored in a register at an adder 13 and restored in the register 14. The cumulative sum over several frames of noise power found from the register 14 by taking the cumulative sum of the signals stored in the register 14 is added to a divider 15 and the mean value of the noise power over the several frames is found by means of the divider 15. On the other hand, a selector 19 selectively takes one of the plural constants set in a constant memory 21 in accordance with the level of the mean value and gives the constant to an adder 22 where the constant is added to the mean value of the noise power. The sum is stored in a register 20 and given to the sound section detector of the next state as a threshold for detecting sound section.



⑨ 日本国特許庁(JP)

⑩ 特許出願公開

⑫ 公開特許公報(A) 平1-244497

⑬ Int. Cl.⁴

識別記号

庁内整理番号

⑭ 公開 平成1年(1989)9月28日

G 10 L 3/00
G 06 F 3/16

3 0 1
3 2 0

A-8842-5D
F-7341-5B

審査請求 未請求 請求項の数 1 (全5頁)

⑮ 発明の名称 音声区間検出回路

⑯ 特 願 昭63-69791

⑰ 出 願 昭63(1988)3月25日

⑱ 発 明 者 杉 伸 夫 神奈川県川崎市幸区柳町70番地 株式会社東芝柳町工場内

⑲ 出 願 人 株 式 会 社 東 芝 神奈川県川崎市幸区堀川町72番地

⑳ 代 理 人 弁 理 士 鈴 江 武 彦 外2名

明 細 書

1. 発明の名称

音 声 区 間 検 出 回 路

2. 特許請求の範囲

入力音声の特徴を抽出する特徴抽出部と、この特徴抽出部で求められた入力音声の特徴情報に従って前記入力音声の音声区間検出に用いる閾値を算出する閾値計算部と、この閾値計算部で求められた閾値に従って前記入力音声の音声区間を検出する音声区間検出部とを備えた音声区間検出回路において、

前記閾値計算部は、音声の入力開始直後における入力信号の短時間パワーの平均値を求める手段と、予め設定された複数の定数の中から上記短時間パワーの平均値に応じた定数を選定する手段と、この選定された定数と前記短時間パワーの平均値との和を閾値として求める手段とを具備したことを特徴とする音声区間検出回路。

3. 発明の詳細な説明

〔発明の目的〕

(産業上の利用分野)

本発明は音声区間検出の為に用いられる閾値を最適設定して高精度な音声区間検出を可能とする音声区間検出回路に関する。

(従来の技術)

近時、マンマシン・インタフェースを実現する1つの手段として音声認識システムが注目されている。この音声認識システムは、基本的には第2図に示すように特徴抽出部1にて入力音声を音響分析してその特徴を抽出する。そして閾値計算部2にて上記特徴抽出部1で求められた特徴の一部を用いて音声区間検出の為に閾値を計算し、更に音声区間検出部3にて上記閾値を用いて入力音声信号のパワーを弁別する等してその音声区間検出がなされる。

その後、検出された音声区間における前記特徴抽出部1で求められた入力音声の種々の特徴をマッチング部4にて、予め標準パターンメモリ5に

登録されている認識対象音声の特徴と照合し、その照合結果を認識部6にて総合判定することで前記入力音声の認識が行なわれる。

尚、マッチング処理は、入力音声の特徴パラメータ系列と認識対象音声の予め求められている特徴パラメータ系列との類似度を複合類似度法を用いてそれぞれ計算する等して行なわれる。そして認識部6において、入力音声の特徴パラメータ系列に対して最も類似度の高い特徴パラメータ系列の認識対象音声のカテゴリを求め、このカテゴリ名を認識結果とすることにより、音声認識処理が行なわれる。

しかしてこのような音声認識の処理過程において、前述した音声区間の検出はマッチング処理に用いられる特徴パラメータ系列自体を左右する重要な役割を担っている。仮に誤った音声区間検出がなされると、その特徴パラメータ系列が異なったものとなるから、誤認識の原因となる。この為、従来より高精度な音声区間検出を実現するべく種々の工夫が試みられている。

- 3 -

め、この平均値を閾値 T_h とすることにより定められる。或いは上記ノイズ・パワーの平均値に予め設定されている微小値（定数）を加算することにより、上記平均値より僅かに高いレベルを閾値 T_h としてることにより設定される。

しかしこのような閾値 T_h の設定の手法にあっては、ノイズ全体のレベルが小さい場合には、定常的には各フレームでのノイズ・パワーがさほど変化することがないので問題を生じることはないが、ノイズ全体が大きい場合、各フレームでのノイズ・パワーが大幅に変化することが希ではない。この為、上述した如く計算されるノイズ・パワーの平均値が、そのときの状態に応じて大幅に変化する虞れがある。この結果、実際に音声区間検出するときのノイズ・パワーとは異なった状態で設定された閾値にて音声区間検出が行なわれ、音声区間検出の精度が大幅に劣化することがあった。

（発明が解決しようとする課題）

このように従来では、音声区間検出の為の閾値を音声入力開始直後の数フレームに亘るノイズ

さて一般的な音声認識システムでは、音声入力信号に対して5~20msec程度の短時間単位（フレーム）毎にその特徴抽出を行なっている。音声区間検出は、通常、その特徴の1つである音声パワーを或る閾値と比較して行なわれる。

しかして一般的には、各フレームにおける音声パワーが或る一定時間以上に亘って上記閾値 T_h を上回ったとき、音声パワーが最初に上記閾値 T_h を上回った時点を音声区間の始点 S として検出している。その後、音声パワーが或る一定時間以上に亘って閾値 T_h を下回ったとき、音声パワーが最初に閾値 T_h を下回った時点を音声区間の終点 E として検出する。そしてこの始点 S と終点 E の区間を音声区間とし、その音声区間の特徴情報を抽出して前述したマッチング処理に供している。

ここで上記閾値 T_h は、通常、音声区間ではない（無音声区間）における入力信号（ノイズ）のパワーを基に、例えば音声の入力を開始した直後の数フレームに亘るノイズ・パワーの平均値を求

- 4 -

パワーの平均値として、或いはその平均値に微小な定数を加算して設定しているだけなので、ノイズ全体が大きく、各フレーム毎のノイズパワーが大幅に変動するような場合には、上述した如き設定される閾値では音声区間を高精度に切出すことができないと云う問題があった。

本発明はこのような事情を考慮してなされたもので、その目的とするところは、ノイズの大きさに拘りなく、そのノイズの大きさに応じた適切な閾値を設定して入力音声の音声区間を高精度に検出することを可能とする実用性の高い音声区間検出回路を提供することにある。

〔発明の構成〕

（課題を解決するための手段）

本発明は入力音声の特徴に従って前記入力音声の音声区間検出に用いる閾値を算出し、この閾値に従って前記入力音声の音声区間を検出する音声区間検出回路において、

音声の入力開始直後における入力信号の短時間パワーの平均値を求めると共に、予め設定された

- 6 -

- 5 -

複数の定数の中から上記短時間パワーの平均値に応じた定数を選定し、この選定された定数と前記短時間パワーの平均値との和を求めて前記音声区間検出に用いる閾値とすることを特徴とするものである。

つまりノイズの大きさを幾つかのレベルに分類し、そのレベルに応じた定数を選定してノイズ・パワーの平均値に加算して音声区間検出の為の閾値を設定することを特徴とするものである。

(作用)

本発明によれば、ノイズ・パワーの平均値にノイズの大きさに応じた定数を加算して閾値が設定されるので、ノイズ全体が大きく、各フレームでのノイズの大きさが大幅に変動する場合には、そのノイズの変動を見込む大きな定数を選定して閾値を設定し、またノイズ全体が小さく、各フレームでのノイズの大きさがほぼ安定しているような場合には小さな定数を選定して閾値を設定することが可能となるので、ノイズの大きさに応じた適切な閾値を設定して入力音声の音声区間検出を

- 7 -

報)は閾値計算部2をパスして音声区間検出部3に直接与えられる。

しかしてスイッチ11を介して取込まれる入力信号は加算器13を介してレジスタ14に格納されている信号と加算され、該レジスタ14に再格納される。この加算器13とレジスタ14は特徴抽出等のフレームに同期して動作し、レジスタ14に格納された信号の加算器13へのフィードバックにより、数フレームに互る入力信号(音声入力開始直後のノイズパワー)を累積加算する。この累積加算によってレジスタ14に求められるノイズパワーの数フレームに互る累積加算値が除算器15に与えられ、そのフレーム数にて除算される。この結果、除算器15から上記ノイズパワーの数フレームに互る平均値Pが求められる。

一方、境界値格納メモリ16にはノイズパワーの平均値Pがどの程度のレベルであるかを識別する為の複数の境界値 $t_1, t_2, \sim t_k$ が設定されている。比較器17はこの境界値格納メモリ16からセレクト18を介して上記境界値 $t_1, t_2, \sim t_k$ を選択

高精度に行なうことが可能となる。

(実施例)

以下、図面を参照して本発明の一実施例につき説明する。

第1図は本発明の一実施例に係る音声区間検出回路における特徴的な部分である閾値計算部の回路構成を示すものである。この実施例回路は第2図に示した従来回路と、基本的には同じ構成を有するものであるが、特徴抽出部1で求められた入力音声の特徴情報に基づいて音声区間検出に用いられる閾値 T_h を計算する閾値計算部2を第1図に示すように構成している点を従来回路と異にしている。

さてこの閾値計算部2は、第1図に示すようにスイッチ11, 12を介して音声入力開始直後の数フレームに互ってその入力信号(特徴抽出部1で求められた特徴情報; 信号パワー)を取込んで動作する如く構成される。この音声入力開始直後の数フレーム以降は、前記スイッチ11, 12の切替によって入力信号(特徴抽出部1で求められた特徴情

- 8 -

的に読出し、前記ノイズパワーの平均値Pと比較している。そしてその比較結果に応じて上記セレクト18を制御し、境界値格納メモリ16から読出す境界値 $t_1, t_2, \sim t_k$ を選択制御し、ノイズパワーの平均値Pのレベルに応じた境界値 t_i を求めている。

この境界値 t_i を求める為の制御信号は以下に説明するセレクト19およびレジスタ20に同時に与えられる。

セレクト19は定数格納メモリ21に設定されている複数の定数 $a_1, a_2, \sim a_{k+1}$ を選択的に求め、これを加算器22に与えて前記ノイズパワーの平均値Pに加算するものである。このノイズパワーの平均値Pに前記定数格納メモリ21から選択的に読出された定数 a_1 を加算した値がレジスタ20に格納され、音声区間検出用の閾値 T_h として次段の音声区間検出部3に与えられる。

ここで上記定数格納メモリ21に格納される複数の定数 $a_1, a_2, \sim a_{k+1}$ は、前記境界値格納メモリ16に格納された複数の境界値 $t_1, t_2, \sim t_k$ に

- 10 -

対応するものである。換言すれば上記境界値 $t_1, t_2, \sim t_k$ はノイズパワーの平均値 P のレベルを $(k+1)$ 個の範囲に区分し、そのレベル範囲毎に設定された $(k+1)$ 個の定数 $a_1, a_2, \sim a_{k+1}$ を選択する為の境界として設定されている。

このように境界値 $t_1, t_2, \sim t_k$ に対応して設定されている定数 $a_1, a_2, \sim a_{k+1}$ が前記セクタ18に連動動作するセクタ19により、ノイズパワーの平均値 P に応じて選択的に求められる。そして前記加算器22にてノイズパワーの平均値 P に加算され、閾値 T_h として設定されることになる。

尚、セクタ18, 19の制御については、例えば先ず最小の境界値 t_1 とノイズパワーの平均値 P との比較を行ない、ノイズパワーの平均値 P が高い場合には次の境界値 t_2 との比較を行なう。それでもノイズパワーの平均値 P が高い場合には更に次の境界値 t_3 との比較を行なうようにし、これを繰返してノイズパワーの平均値 P が境界値 t_1 より下回った時点でその選択動作を停止制御

— 1 1 —

する等して行なわれる。勿論、レベルの高い境界値側から逆向きにその制御を行なうことも可能であり、中心の境界値から比較処理を進めることも可能である。

かくしてこのように構成された本回路によれば、ノイズパワーのレベルに応じた定数を選定してそのときのノイズパワーの平均値 P に加算し、音声区間検出の為の閾値 T_h を設定するので、ノイズパワーのレベルに応じた適切な閾値 T_h の下で音声区間検出を高精度に行なうことが可能となる。しかも上述したように定数の選定制御を非常に簡易に行なうことができ、処理負担の増大を招くこともない等の効果が奏せられ、実用的利点が多大である。

尚、本発明は上述した実施例に限定されるものではない。例えばノイズパワーの平均値のレベルを分解する境界値の数や、各レベル範囲に対して設定する定数の値はそのシステム仕様等に応じて定めれば良いものである。またノイズパワーを求める為のフレーム区間についてもそのときの音声

— 1 2 —

区間に掛からない範囲で定めれば良いものである。その他、本発明はその要旨を逸脱しない範囲で種々変形して実施することができる。

〔発明の効果〕

以上説明したように本発明によれば、非常に簡易にしてノイズレベルに応じた適切な閾値を設定し、音声区間検出を高精度に行ない得る等の実用上多大なる効果が奏せられる。

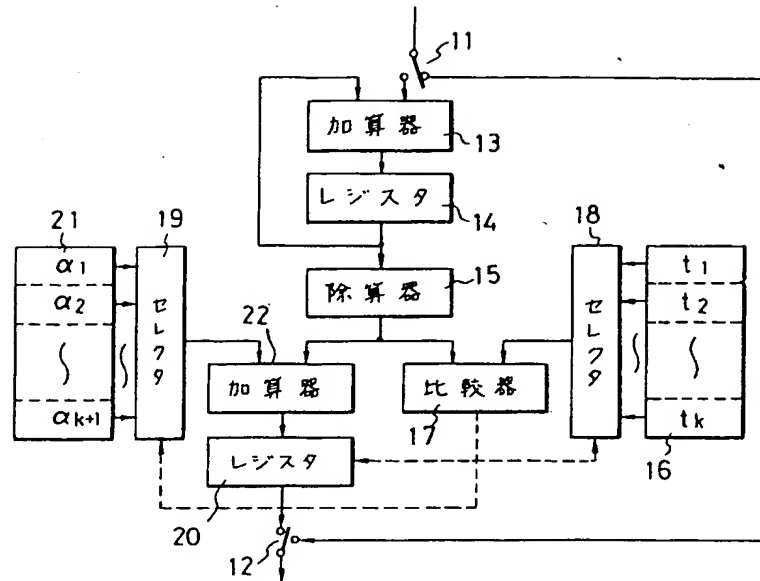
4. 図面の簡単な説明

第1図は本発明の一実施例に係る音声区間検出回路における閾値計算部の回路構成図、第2図は音声認識システムの基本的な構成例を示す図である。

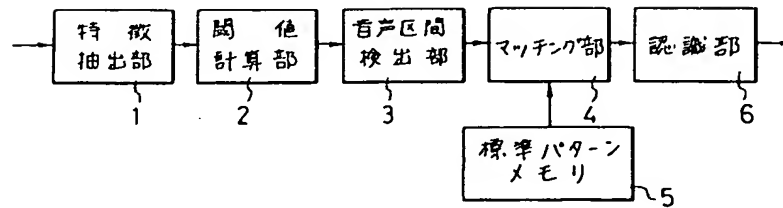
1…特徴抽出部、2…閾値計算部、3…音声区間検出部、13, 22…加算器、14, 20…レジスタ、15…除算器、16…境界値格納メモリ、17…比較器、18, 19…セクタ、21…定数格納メモリ。

出願人代理人 弁理士 鈴江武彦

— 1 3 —



第 1 図



第 2 図